Ashraf *et al. Journal of Electrical Systems and Information Technology* (2026) 13:37
https://doi.org/10.1186/s43067-026-00340-2

Journal of Electrical Systems and Information Technology

RESEARCH Open Access

Decomposition-based multi-modal image fusion for breast cancer classification using AlexNet and MCFO filter

Basem Ashraf[1], El-Sayed M. El-Rabaie[2] and Nariman Abdel-Salam[3*]

*Correspondence:
Nariman Abdel-Salam
nariman.abdelsalam2020@gmail.com
[1]Department of Communications and Electronics, Faculty of Engineering, Canadian International College (CIC), Shiekh Zayed City, Egypt
[2]Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt
[3]Electrical Communication and Electronic Systems Engineering Department, Faculty of Engineering, October University for Modern Sciences and Arts (MSA), 6th of October City, Egypt

Abstract

Breast cancer is one of the most common and deadly cancers, affecting millions worldwide. Early and accurate detection is essential for effective treatment and improved patient outcomes. Advances in medical imaging technologies, such as Digital Mammography (DM), Ultrasound (US), and Magnetic Resonance Imaging (MRI) provide clinicians with detailed information about breast tumors and surrounding tissues. However, merging and analyzing these multimodal images pose challenges. Medical image fusion combines images from different modalities to improve quality, reduce noise and redundancy, and support more precise clinical decisions. In this study, three models were developed to evaluate feature extraction strategies: Model A uses an AlexNet architecture, Model B employs a LeNet-5 architecture, and Model C incorporates a DenseNet-121 architecture. All models are integrated with a decomposition method, such as PCA or DWT, for image fusion into three main categories: normal, benign, or malignant. The Modified Central Forced Optimization (MCFO) filter is employed to enhance diagnostic accuracy. Our framework was tested on a new dataset from Baheya Hospital in Egypt, which includes high-quality, annotated images. Results show that combining DWT-based methods with AlexNet and the MCFO filter achieves top performance, with an accuracy of 97.4%, a precision of 95%, a Recall of 96%, a F1 Score of 93%, and an ROC score of 96.95%, with minimal loss, demonstrating strong generalization and stability across epochs. These findings highlight the superior performance of the DWT-based approach with AlexNet and MCFO compared to other methods.

Keywords Breast cancer, Multimodal image fusion, Principal Component Analysis (PCA), Discrete Wavelet Transform (DWT), AlexNet, LeNet, DenseNet, Modified Central Forced Optimization Filter (MCFO)

Introduction

Worldwide, breast cancer is the leading cause of cancer-related death among women ages 20 to 60 [1]. The World Health Organization [2] estimates that in 2020, 2.3 million women will be diagnosed with breast cancer, with 685,000 dying from the disease. Malignant cells form abnormal masses within tissues. These cells spread throughout the

body and into blood vessels. They travel through the body, damaging other tissues and organs. Recent studies show that breast cancer accounts for 29% of all newly diagnosed cancer cases in women. Men can also develop breast cancer, although it is uncommon [3, 4]. When caught early, the survival rate can rise to 80%. Various factors, including age and family history, influence the risk of developing breast cancer. Breast tissue is categorized into three main types [5]:

- Normal: Healthy breast tissue without abnormal cell growth. This category provides a baseline for detecting benign or malignant changes.
- Benign: The tumor is harmless to health if the cells are not malignant. It won't spread to other parts of the body or invade nearby tissues. Unless it damages surrounding tissues, nerves, or blood vessels, a benign tumor is not a cause for concern.
- Malignant: This indicates that the tumor consists of cancerous cells and can spread to nearby tissues, making it dangerous. Metastasis is the process by which cancer cells enter the bloodstream or lymph nodes and spread to other parts of the body. This tumor is more lethal and hazardous.

MRI, US, and DM are common techniques used to detect breast cancer [6, 7]. Radiologists examine breast abnormalities identified by screening DMs [8]. MRI can detect deeply damaged tissue, and ultrasound can determine whether a breast tumor is a liquid cyst or a solid mass [9]. To improve segmentation, pigment and texture features from histopathology images are utilized. Although medical image analysis techniques for histopathology are advancing, automated approaches for cost-effective and efficient diagnosis remain necessary [10, 11].

Image fusion (IF) involves combining images from multiple sensors to create an informative image that improves medical diagnosis and decision-making. Image Registration (PR) is the initial step in the fusion process, where reference images are used to align the source images. Image fusion approaches are categorized into three types: pixel-based, decision-based, and feature-based [12, 13]. When the original images come from the same sensor, (IF) is sometimes called multi-focus, multi-view, or multi-temporal fusion. Multi-focus fusion merges photographs taken with different focus lengths simultaneously. Multi-view fusion combines images captured from different viewpoints at the same time. In medical imaging, images from DM, MRI, and/or US can be combined simultaneously to emphasize differences or create realistic representations of parts that were not captured at the same time. Multi-temporal fusion integrates images taken at different times [14]. The medical field needs the fusion of various multimodal images because each modality alone often offers insufficient information for accurate diagnosis. The fused image allows for more precise diagnoses, such as detecting tumor regions and breast abnormalities [15–18]. Since the fused image incorporates multiple data sources that cannot be obtained from a single image, image fusion is a valuable technique for analyzing and utilizing large volumes of images close to the source [19, 20]. The deep learning architecture was developed to overcome the limitations of traditional machine learning methods by performing the entire task end-to-end and learning the full range of features from low to high levels, especially for breast cancer classification [21–23].

Beyond AlexNet, this study also examines the effectiveness of two other deep convolutional neural network architectures: LeNet-5 and DenseNet-121. LeNet-5 is a convolutional neural network. Despite its relatively shallow structure, with seven layers

of alternating convolutional and subsampling layers, LeNet-5's main ability to detect strokes, edges, and shapes provides a solid foundation for analyzing medical images [24].

In contrast, DenseNet-121 is a more recent deep convolutional neural network architecture that connects each layer directly to all subsequent layers. This dense connection pattern promotes feature reuse, mitigates the vanishing-gradient problem, and greatly reduces the number of parameters compared to older models. DenseNet-121 has shown strong feature extraction abilities that can be applied in medical imaging tasks [25].

This framework enhances the diagnostic classification of breast images (normal, benign, and malignant) by comparing three proposed approaches: AlexNet, LeNet-5, and DenseNet-121 architectures for feature extraction, combined with two decomposition methods for image fusion: PCA and DWT. All approaches use the (MCFO) filter to improve classification accuracy. The proposed methods were evaluated using a new dataset from the Baheya Foundation for Early Detection and Treatment of Breast Cancer in Egypt. A list of abbreviations is provided in Table 1.

The main contributions of this paper are summarized as follows

- Proposal and comparative analysis of three different model architectures for classifying breast cancer multimodal images: AlexNet, LeNet-5, and DenseNet-121.
- Performing a comparative analysis of the fusion of multimodal medical images using two different decomposition methods: PCA and DWT for dimensionality reduction.
- Enhancing the fusion results' performance using the MCFO filter.
- The main advantage of using the AlexNet deep learning architecture for breast cancer diagnosis after image fusion with DWT or PCA is that it can effectively learn complex patterns and features from the fused medical images.
- The proposed methods were evaluated using a new breast cancer database from the Baheya Foundation for Early Detection and Treatment of Breast Cancer in Egypt.

Table 1 List of abbreviations

Abbreviations	Explanation
IF	Image Fusion
DL	Deep Learning
ML	Machine Learning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
MMIF	Multimodal Medical Image Fusion
US	Ultrasound
AE	Autoencoder
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
PCA	Principal Component Analysis
DWT	Discrete Wavelet Transforms
AVG. G	Average Gradient
UIQI	Universal Image Quality Index
SSIM	Structural Similarity Index Measure
ROC	Receiver Operating Characteristic
RMSE	Root Mean Square Error
PSNR	Peak Signal-to-Noise Ratio
ERGAS	Error in Relative Global Dimensionless Synthesis

- The performance of fusion methods was evaluated using six metrics: SSIM, PSNR, RMSE, CC, ERGAS, and UIQI.
- The performance of the deep learning method was evaluated using four metrics: Accuracy, Precision, Recall, F1-Score, and ROC-AUC.
- Improving the diagnostic classification of breast images (normal, benign, and malignant).

The rest of the study is organized as follows: Sect. 2 reviews the literature. Section 3 offers a detailed explanation of the proposed framework, including fusion methods and classification algorithms. Section 4 describes the experimental setup in detail. Section 5 presents the simulation results and compares the findings. Finally, Sect. 6 summarizes the conclusions.

Related work

The fused image provides a more accurate diagnosis, such as locating tumors and breast abnormalities. Since the fused image combines data that can't be obtained from a single source image, image fusion is an effective technique for analyzing and using large volumes of images from the source.

Yusuf et al. [26] proposed an automated method to detect invasive ductal carcinoma (IDC) using 277,524 image patches from a public histopathology dataset. They utilized deep transfer learning with two CNN pre-trained models, ResNet-50 and DenseNet-161, which achieved accuracy scores of 91.57% and 90.96%, respectively.

In addition to comparing six different activation functions, Ibrahim et al. [27] proposed a CNN architecture with 15 convolutional layers and two fully connected layers. The PatchCamelyon dataset included 57,000 unannotated images for testing and 220,000 annotated images for training. The highest AUC achieved by their model was 95.46%.

Eman et al. [28] presented various methods for classifying breast cancer, including magnetic resonance imaging, ultrasound, and mammography. Fusion can be classified as either early or late. Early fusion involves direct connections between modalities, while late fusion focuses on learning the complex relationships among multiple modalities. As a result, late fusion methods generally yield more accurate results. They achieved an F-score of 94.65% and a balanced accuracy of 95.2%.

Othman et al. [29] developed a hybrid deep-learning model that makes decisions using data from multiple sources by designing and implementing two different classifiers. Compared to predictions based on a single data modality, incorporating multi-omics data (clinical, gene expression, and copy number alteration data) from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset is expected to improve the accuracy of patient survival predictions. They achieved an accuracy of 97.0%.

A feature and decision fusion approach for diagnosing breast cancer was reported by Yadav, Rohit, et al. [30]. Low-contrast mammography images were initially enhanced using CLAHE technology. They then applied a CNN to extract features. Classifiers were developed with machine learning methods (SVM, DT, and RF). The ensemble voting classifier produced the final prediction. The accuracy rates for the three algorithms were 95%, 94.3%, and 92.3%. The results indicated that RF outperformed SVM and DT in BC prediction. Overall accuracy increased to 96.18% when using voting classifiers (SVM, DT, and RF).

To assist with breast cancer diagnosis and classification, Sahar et al. [31] compared the pre-trained model AlexNet with a proposed machine learning-based model for breast cancer detection and classification. They analyzed three datasets- A, B, and C- and combined them to create two new datasets, A2C and B3C. Dataset A2C is a fusion of A, B, and C with two classes: benign and malignant, whereas Dataset B3C merges A, B, and C into three classes: benign, malignant, and normal. Dataset A contains 780 images collected from [32, 33], Dataset B has 7783 images collected from [33, 34], and Dataset C includes 1126 images collected from [32]. They achieved an accuracy of 94.5% on Dataset B3C and 94.9% on Dataset A2C.

Hamdi et al. [35] demonstrated that combining data from multiple image modalities, such as mammography and ultrasound, can significantly improve classification accuracy. Since they allow deep supervision and effective gradient flow during training, dense connections are commonly used in computer vision. DenseNet achieves impressive results in natural image classification by using feed-forward connections between all layers. It is possible to learn more complex combinations of modalities with DenseNet 201, which connects similar and different routes. Multi-modal images gather features from various perspectives and provide additional information. The reported metrics were 93.83% for accuracy, recall, and precision; 95.61% for the area under the curve; and 93.8% for the F1 score.

Hussain et al. [36] presented a multimodal dataset that includes imaging and textual data, such as clinical and radiological features. They proposed a DL-based Multiview Multimodal Feature Fusion (MMFF) strategy for breast cancer classification, which uses images and tabular data from an in-house dataset they created. Imaging features were extracted using a ResNet50 model, while textual features were obtained with an artificial neural network (ANN). Afterwards, features from both modalities were fused using a late feature fusion strategy. The model achieved an AUC of 0.874.

Mahmood et al. [37] introduced a Multi-Modal Feature Fusion Network for Histopathology (MFF-HistoNet) to address the challenges of multi-grading breast images and to significantly enhance diagnostic accuracy. MFF-HistoNet combines a CNN with a Quantum Tensor Network (QTN). The GLCM method is used alongside LBP and Gabor filtering to extract local cell shape features of histopathological images across space, scales, and orientations. The MFF-HistoNet al.gorithm achieved an accuracy of 98.8% at the image level and 98.4% at the patient level under 100× magnification, and 98.1% and 98.9% under 40× magnification.

Ghantasala et al. [38] introduced HXM-Net, a deep learning model specifically designed to improve breast cancer detection by combining multi-modal ultrasound imaging. HXM-Net uses Convolutional Neural Networks (CNNs) to extract spatial features and Transformer-based fusion to effectively combine information from B-mode and Doppler ultrasound images. HXM-Net achieved an accuracy of 94.20%, a sensitivity (recall) of 92.80%, a specificity of 95.70%, an F1 score of 91.00%, and an AUC-ROC of 0.97%.

Recent studies show that many scenarios can improve complementary information through well-designed deep learning and feature fusion modules. However, it is important to highlight the mishandling of the following, as shown in Table 2.

- Feature uncorrelation can reduce fusion performance in specific cases.
- A limited number of medical images are available, so more preprocessing is needed.

Table 2 Literature of previous studies

Ref., Year	Detection techniques	Models	Fusion strategies	Dataset	Results
[26], 2020	ResNet-50 and DenseNet-161	Invasive ductal carcinoma (IDC)	N. A.	The public histopathology dataset contains 277,524 images.	This has yielded an F-score of 92.38% and a balanced accuracy value of 91.57%.
[27], 2020	Convolutional neural network (CNN)	Classification of histopathology images	N. A.	Patch Camelyon	AUC of 95.46%.
[28], 2022	DenseNet 201	Mammography, magnetic resonance, and ultrasound	Earlier fusion and later fusion	Mini-DDSM And BUSI	F-score of 94.65% and balanced accuracy value of 95.2%
[29], 2023	(CNN) Architecture	The gene expression profile data	LSTM and GRU	METABRIC dataset	The accuracy achieved by LSTM is 97.0%.
[30], 2022	(CNN) Architecture	DMs	Decision fusion (SVM, decision tree, and random forest)	MIAS dataset	The accuracy achieved by SVM is 92.30, while using decision fusion (SVM, decision tree, and random forest), the highest accuracy of 96.12% is achieved.
[31], 2023	AlexNet	Ultrasound and histopathological images.	Fuzzing these datasets and got 2 datasets, A2C and B3C.	Three customized datasets, A, B, and C.	Accuracy of 94.5% on Dataset B3C and 94.9% on Dataset A2C
[35], 2021	DenseNet-201	Mammography and ultrasound	Early and late fusion	Mini-DDSM And BUSI	Accuracy, recall, precision, area under the curve, and F1 score were 93.83%, 93.83%, 93.83%, 95.61%, and 93.8%, respectively.
[36], 2024	Efficient-Netb7 + ANN	Mammography	Multiview multimodal feature fusion (MMFF)	Developed an in-house dataset	AUC of 0.874
[37], 2025	MFF-HistoNet combines a CNN and a Quantum Tensor Network (QTN).	Histopathological images	The GLCM method is fused with LBP and Gabor filtering.	The BreaKHis dataset	Accuracy of 98.8% at the image level and 98.4% at the patient level under 100× magnification.
[38], 2025	HXM-Net	Multi-modal ultrasound imaging	Transformer-based fusion	Developed an in-house dataset	Accuracy of 94.20%, sensitivity (recall) of 92.80%, specificity of 95.70%, an F1 score of 91.00%, and AUC-ROC of 0.97

- Small, unbalanced datasets.
- Some imaging modalities are unavailable.
- Limited use of diverse models in digital mammograms.

To overcome the limitations of current image fusion techniques, this paper proposes a superior decomposition-based multi-modal image fusion method for breast cancer classification.

Proposed methodology

System overview

Figure 1 presents a general flow diagram of the proposed system. The goal is to classify breast cancer using a multi-modal image fusion technique. This method combines

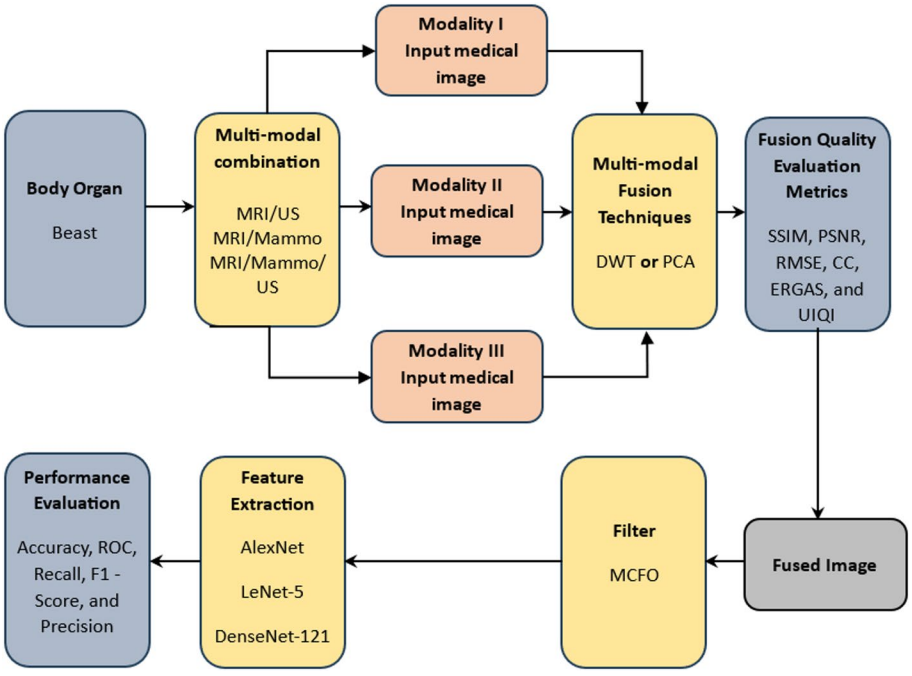


Fig. 1 The block diagram of the general framework

medical images from breast cancer imaging datasets. The datasets used are customized from the Baheya Foundation for Early Detection and Treatment of Breast Cancer, featuring three different modalities: ultrasound, DM, and magnetic resonance imaging. Image fusion is based on Principal Component Analysis (PCA) or Discrete Wavelet Transform (DWT). The fused images are enhanced using a modified central forced optimization (MCFO) filter. For classification, key features are extracted using deep learning architectures such as AlexNet, LeNet-5, and DenseNet-121. Finally, the system's performance is assessed with various evaluation measures, detailed in the sections below. The goal is to improve diagnostic capabilities by integrating data from multiple modalities and databases.

Medical imaging types

Image processing techniques, such as image fusion, are becoming increasingly important in modern medicine and healthcare by deriving clinical data from multimodal medical imaging, including MRI-MRA, MRI-PET, CT-MRI, MRI-SPECT, PET-CT, X-ray-US, US-MRI, and US-DM. This paper initially analyzes three types of medical images for breast cancer: MRI, DM, and US [3].

Magnetic resonance imaging (MRI)

It combines a magnetic field and a radio wave to produce precise images of breast tissue; we use it as a functional modality to show blood flow and verify each tissue's proper function. This medical imaging technique employs magnetic flux, radiofrequency related to the illness, and various biological processes. The main feature of MRI is that it creates "slices" of the human body using magnetic signals, revealing details about damaged soft tissues. Since "multi-model imaging" involves generating signals for multiple imaging techniques simultaneously, image fusion overcomes the limitations of multi-modal

imaging by enabling the recovery and prediction of missing MRI data. Thus, by combining MRI with other images, missing data can be restored. When used with other modalities, MR images tend to improve clinical usefulness and imaging accuracy with current image fusion technology. Several researchers have worked on integrating MRI with other modalities using image fusion techniques. Common pairings include MRI-PET and CT-MRI [39].

Ultrasound imaging (US)

Ultrasound is commonly used alongside mammography and utilizes sound waves to produce images of breast tissue. It functions as a structural technique to describe breast anatomy and tumor boundaries. The main mechanism of US imaging is the slight vibrations generated by the radiation energy within the body. To enhance breast cancer diagnosis, vibro-acoustography (VA), another imaging method, combines mammography with ultrasound-animated acoustic emission. US imaging is also being researched as a non-invasive method to harvest tissues from the liver, breast, prostate, and arteries. It can evaluate variations in mechanical response to resonance in specific environments, serving as a nondestructive technique to detect hidden flaws in materials. Since mammography X-rays cannot reveal the length and density of structures, ultrasonic imaging remains valuable, especially considering tissue thickness. Applications include analyzing breast development and detecting mass lesions. Furthermore, a fusion strategy based on pixels or color can provide more detailed and visual information by combining images from two different sensors, X-ray and ultrasound, thus improving diagnostic data [39].

Digital mammography (DM)

For screening and diagnosis, a personalized breast X-ray image is used. DM is the most effective tool for detecting and diagnosing breast cancer early. Each breast is shown in two images: medial-lateral-oblique (MLO) and craniocaudal (CC). DM offers benefits such as a lower death rate, earlier treatment, and safe radiation exposure.

Image fusion techniques

Data preprocessing was conducted to enhance data quality and create a clean dataset suitable for model development. Without preprocessing, issues such as inconsistencies, errors, noise, missing data, and model overfitting can arise. To increase accuracy and quality while maintaining the complementary information of the images, the multimodal image fusion approach combines multiple images from one or more imaging modalities [6]. The most common medical image fusion methods include CT, MRI, US, and SPECT. Although SPECT and PET scans have lower spatial resolution, they provide valuable information about the body, including blood flow, soft tissue movement, and metabolism. As shown in Fig. 2, high-resolution images produced by MRI, CT, and US scans offer detailed anatomical information. Usually, multimodal images are created by merging structural and functional images to provide additional insights that assist medical professionals in diagnosing clinical conditions.

Principal component analysis (PCA)

It is a statistical technique that retains important information while reducing the data's dimensionality. The fusion process is expected to include extracting key elements from

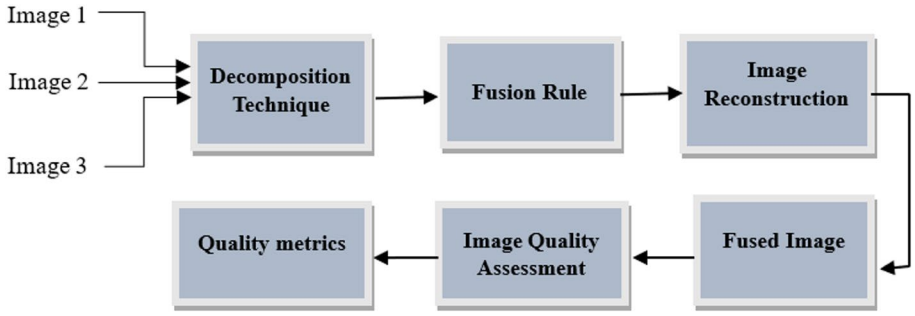


Fig. 2 **A** block diagram describing the fusion steps

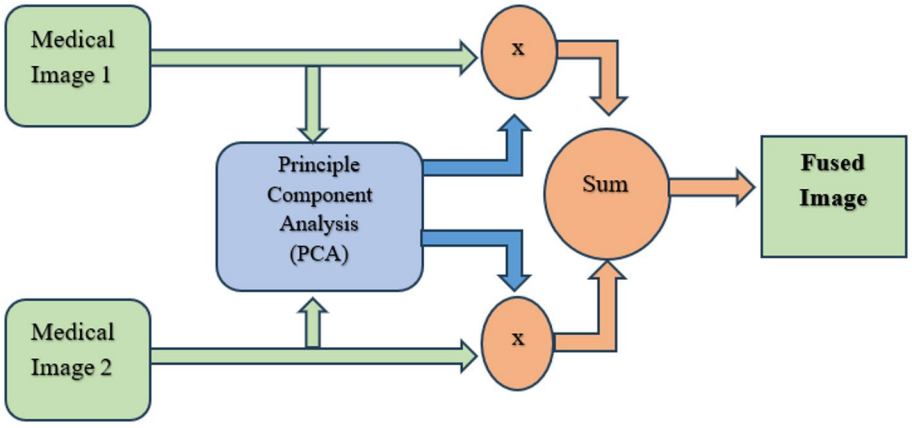


Fig. 3 Medical image fusion scheme using PCA transform

each modality and merging them to create a comprehensive view of breast tissue. PCA is a powerful dimensionality reduction technique that can effectively combine medical images from multiple modalities into a single fused image representation [7]. The fused image captures the most relevant information from the input images, providing a complete perspective of tumor characteristics, as shown in Fig. 3. The key steps in this process are:

1. Data Preprocessing: Collect a set of medical images (DMs, ultrasounds, and MRIs) from breast cancer patients. Preprocess the images by normalizing pixel values, resizing them to a uniform size, and converting them to a standard format (e.g., grayscale).
2. Feature Extraction: Apply PCA to the preprocessed images to extract the most important features that capture the underlying patterns in the data. PCA identifies the principal components (PC) that account for the maximum variance in the data, effectively reducing the dimensionality of the input images.
3. Image Fusion: Combine collected PCs from various imaging modalities to create a single image representation. This fusion can be achieved by concatenating the PCs from each modality or by using weighted averaging, depending on the importance of each PC.
4. Classification and Evaluation: Deep learning classifiers can use the fused image representation to diagnose and stage breast cancer. The classifier may be trained on fused images to identify patterns associated with different stages or types of breast

cancer. Breast cancer detection and staging methods can be evaluated using relevant metrics such as accuracy, sensitivity, and specificity. Validation with an independent dataset is essential to confirm the generality of the technique. The fused images are then evaluated using "Evaluation Metrics" to assess the quality of the fusion process. Metrics like the signal-to-noise ratio, structural similarity index, or other performance indicators relevant to image fusion tasks may be employed as evaluation tools.

Discrete wavelet transform (DWT)

This technique is used to merge two or more images to create a higher-quality final image. Let's explain the function of each block in this diagram, as shown in Fig. 4.

(1) DWT (Discrete Wavelet Transform): This block applies the discrete wavelet transform to the source image, breaking it down into various frequency components [8]. The output consists of wavelet coefficient maps that represent the spectral and spatial information of the source image.
(2) Fusion Decision: This step determines how to combine the wavelet coefficient maps obtained from multiple source images. It employs various fusion algorithms to select the best values from each wavelet coefficient map to generate the fused wavelet coefficient maps.
(3) Fused Wavelet Coefficient Maps: This section presents the fused wavelet coefficient maps produced by the fusion decision block. These maps contain combined spectral and spatial information from multiple source images.
(4) Inverse DWT (IDWT): This section performs the inverse discrete wavelet transform on the fused wavelet coefficient maps. The result is the final fused image, which integrates both spectral and spatial information from multiple source images. This diagram generally shows how the discrete wavelet transform technique is used to combine several source images and generate a final image of higher quality than the original ones.

Modified central forced optimization (MCFO) filter

The MCFO filter is a variation of the central forced optimization (CFO) filter, a nonlinear method used for image denoising. It is designed to overcome the limitations of the original CFO filter, which can sometimes create artifacts or fail to preserve important

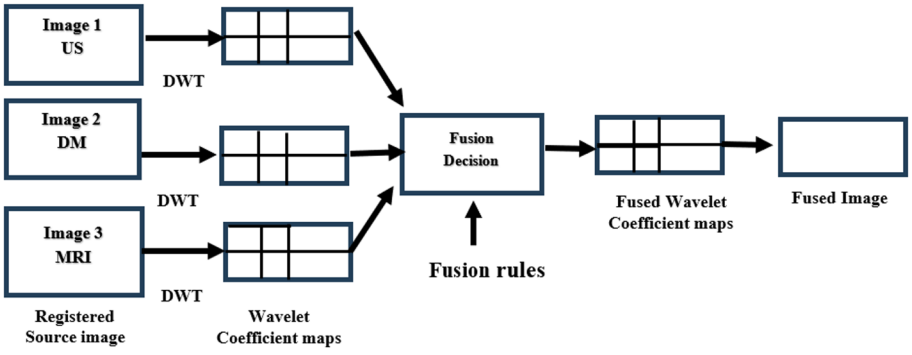


Fig. 4 Medical image fusion scheme using wavelet transform

image details. The MCFO filter adds extra steps to enhance the CFO filter's performance, such as adaptive window size selection and an improved cost function optimization. It is especially effective at removing Gaussian, impulse, and mixed noise from images while maintaining edges and other key features. In various image processing applications, the MCFO filter outperforms other advanced denoising methods, including median and Wiener filters. Figure 5 demonstrates that the MCFO filter is computationally efficient and easy to apply in real-time scenarios, making it suitable for many video and image processing tasks [10]. The main steps of the modified central forced optimization (MCFO) filter are as follows:

1. Position Feedback Information: This indicates the current position of the quadrotor, which is measured and sent back to the control system. Position feedback can come from sensors like GPS, vision-based systems, or other tracking methods. This feedback is used to compare the actual position of the quadrotor with the desired position.
2. Path Planning Process: This section generates a desired, time-varying position path that the quadrotor should follow. The path-planning process may involve algorithms that consider factors such as obstacles, waypoints, and optimization criteria to produce a smooth, feasible trajectory. The output of this block is the desired position path over time, represented by position coordinates.
3. Position Controller (Outer Loop): This is the outer control loop that manages the quadrotor's position. It uses position feedback and the desired position path as inputs. The position controller compares the actual position with the desired position and produces the expected time-varying attitude information (Euler angles) needed to follow the path. These Euler angles represent the desired orientation of the quadrotor, which the inner attitude controller will aim to achieve.
4. Attitude Controller (Inner Loop): This is the inner control loop that manages the attitude (orientation) of the quadrotor. It receives the expected time-varying attitude information (Euler angles) from the position controller as the desired attitude. The attitude controller uses the current attitude feedback and the desired attitude to generate the appropriate control signals. These signals are then sent to the quadrotor's dynamics model to activate the motors and achieve the desired orientation.
5. Dynamics Model of the Quadrotor Helicopter: This component represents the mathematical model that describes the quadrotor helicopter's dynamics. It includes the equations of motion, aerodynamic forces, and other physical properties of the quadrotor. The control signals from the attitude controller are applied to this model, which then updates the state of the quadrotor (position, velocity, orientation, etc.).

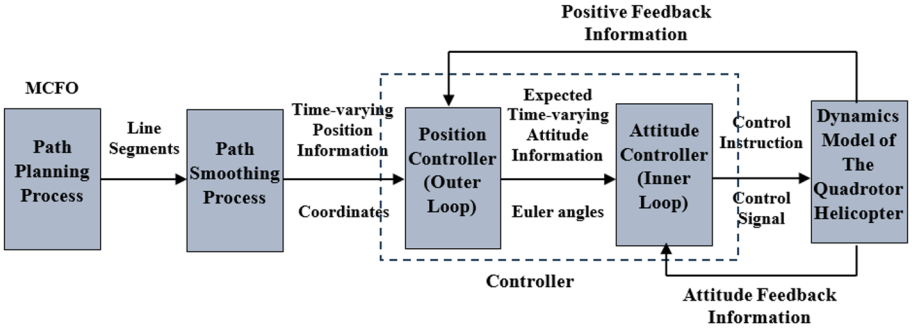


Fig. 5 Main steps of the modified central forced optimization (MCFO) filter

6. Attitude Feedback: This indicates the current attitude (orientation) of the quadrotor. This information can be obtained from sensors like gyroscopes, accelerometers, or vision-based systems. This feedback is used by the attitude controller to compare the actual orientation with the desired one and to make the necessary adjustments.

The MCFO filter is a nonlinear optimization-based denoising method designed to improve fused medical images while maintaining edges and clinical features. Given an input fused image, I(x, y), of size M × N, the MCFO filter works by minimizing a cost function that balances fidelity to the original image and smoothness constrained by local statistics.

The MCFO is a population-based metaheuristic algorithm that explores a Decision Space (DS) by flying a group of probes (Np), with trajectories governed by equations similar to the laws of gravitational motion in the physical universe [9]. This technique primarily has three parameters for each probe: position vector (R), acceleration vector (A), and fitness value (M). The MCFO has two main modifications over the CFO to enhance its accuracy and memory capability for updating probe positions, causing them to be attracted to the best previously visited positions according to the following Eq.

$$\boldsymbol{A}_{j-1}^{\boldsymbol{p}} = \boldsymbol{G}_{\boldsymbol{j}} \sum_{\boldsymbol{k}=1}^{\boldsymbol{N}_{\boldsymbol{p}}} \boldsymbol{U}\left(\boldsymbol{M}_{\boldsymbol{j}-1}^{\boldsymbol{k}} - \boldsymbol{M}_{\boldsymbol{j}-1}^{\boldsymbol{p}}\right) \times \left(\boldsymbol{M}_{\boldsymbol{j}-1}^{\boldsymbol{k}} - \boldsymbol{M}_{\boldsymbol{j}-1}^{\boldsymbol{p}} \frac{\boldsymbol{\alpha}\left(\boldsymbol{R}_{\boldsymbol{j}-1}^{\boldsymbol{k}} - \boldsymbol{R}_{\boldsymbol{j}-1}^{\boldsymbol{p}}\right)}{\| \boldsymbol{R}_{\boldsymbol{j}-1}^{\boldsymbol{k}} - \boldsymbol{R}_{\boldsymbol{j}-1}^{\boldsymbol{p}} \|}\right) \tag{1}$$

$$\boldsymbol{G}_{\boldsymbol{j}} = \boldsymbol{G}_0 \mathbf{exp}\left(-\frac{\boldsymbol{j}\boldsymbol{y}}{\boldsymbol{N}_{\boldsymbol{t}}}\right) \tag{2}$$

$$\boldsymbol{R}_{\boldsymbol{j}}^{\boldsymbol{p}} = \boldsymbol{R}_{\boldsymbol{j}-1}^{\boldsymbol{p}} + \boldsymbol{C}_{1,\boldsymbol{j}}\, \boldsymbol{r}\mathbf{tan}\boldsymbol{\theta}\ \boldsymbol{A}_{\boldsymbol{j}-1}^{\boldsymbol{p}} \boldsymbol{\Delta}\, \boldsymbol{t}^2 + \boldsymbol{C}_{2,\boldsymbol{j}}\, \boldsymbol{r}\mathbf{tan}\boldsymbol{\theta}\ \left(\boldsymbol{R}_{\mathbf{best}} - \boldsymbol{R}_{\boldsymbol{j}-1}^{\boldsymbol{p}}\right) \boldsymbol{\Delta}\, \boldsymbol{t},\ \boldsymbol{j} \geq 1 \tag{3}$$

$$\boldsymbol{C}_{1,\boldsymbol{j}} = \boldsymbol{C}_1^{\mathbf{max}} - \left(\frac{\boldsymbol{C}_1^{\mathbf{max}} - \boldsymbol{C}_1^{\mathbf{min}}}{\boldsymbol{N}_{\boldsymbol{t}}}\right) \times\ \boldsymbol{j} \tag{4}$$

$$\boldsymbol{C}_{2\boldsymbol{j}} = \boldsymbol{C}_2^{\min} - \frac{\boldsymbol{C}_2^{\max} - \boldsymbol{C}_2^{\min}}{\boldsymbol{N}_{\boldsymbol{t}}} \times\ \boldsymbol{j} \tag{5}$$

where G_j is the value of the current gravitational constant, G_o is the initial gravitational constant, γ is the descent coefficient factor, p is the probe number, N_t is the maximum number of iterations, C_1 and C_2 are the time-varying acceleration coefficients, $rand_1$ and $rand_2$ are two random numbers in the range [0, 1], U(.) is the unit step function, α and β are the CFO exponents, and Δt is taken as a unit time step increment. For the proposed medical image fusion algorithm, the fitness value is based on maximum local contrast, maximum entropy, or optimal PSNR of the fused images. These metrics were chosen because they are the most commonly used and trusted measures for evaluating image quality. The gain parameter values a1, b1 for high-pass sub-bands and a_2, b_2 for low-pass sub-bands lie within the interval [0–1], under the constraint $a_1 + b_1 = 1$, and $a_2 + b_2 = 1$.

Deep learning methods

AlexNet architecture

Having been trained on more than a million images, AlexNet is a pre-trained deep convolutional neural network (DCNN) that can classify images into 1, 000 different categories. This network accepts a 227 × 227 × 3 image as input and consists of eight layers [11]. The first convolutional layer filters the input image using 96 kernels (11 × 11 × 3) with a stride of 4 pixels. The second layer contains 256 kernels (5 × 5 × 48). The first two convolutional layers are followed by a max- activation pooling layer and a normalization layer. The third, fourth, and fifth convolutional layers are not separated by pooling layers. The outputs of the second convolutional layer are connected to 384 kernels of size 3 × 3 × 256 through the third convolutional layer. The fourth convolutional layer has 384 kernels (3 × 3 × 192), and the fifth has 256 kernels (3 × 3 × 192). After the fifth convolutional layer, a max-pooling layer outputs into two fully connected layers, each containing 4096 neurons. The SoftMax classifier provides 1, 1000 class labels from the second fully connected layer. A dropout layer with a rate of 0. 5% is applied in the first two fully connected layers, and each of the first seven layers uses a Rectified Linear Unit (ReLU) activation function. This study fine- tunes the last layers of the original AlexNet architecture to address the breast mass classification problem in DM images. The feature extraction layers of the original model are preserved, while the final three layers are replaced with new layers: a fully connected layer, a SoftMax layer, and a classification output layer specifically designed for the new dataset. The number of outputs in the final layer is reduced from 1,000 to three, corresponding to the classes of benign, normal, and malignant. The fully connected layers are set to have three hidden layers to accommodate this, as shown in Fig. 6. The model classifies images into three categories rather than the typical 1,000 classes for which AlexNet was originally designed. The AlexNet architecture functions as an end- to- end classifier using SoftMax.

LeNet-5 architecture

Being one of the earliest convolutional neural networks, LeNet-5 is a classic DCNN architecture. The standard LeNet-5 architecture, as shown in Fig. 7, is relatively shallow compared to modern networks. It typically accepts a 32 × 32 × 1 grayscale image as input and contains seven layers, including convolutional and subsampling layers. The network uses 6 kernels of size 5 × 5 in the first convolutional layer to extract low-level features like edges and curves. This is followed by a subsampling, or pooling, layer to reduce the spatial dimensions. A second convolutional layer applies 16 larger kernels of size 5 × 5 × 6 to learn more complex features [12]. This is followed by another subsampling layer. The output is then flattened and passed through a fully connected convolutional layer, often referred to as C5, with 120 feature maps, followed by a fully connected layer called

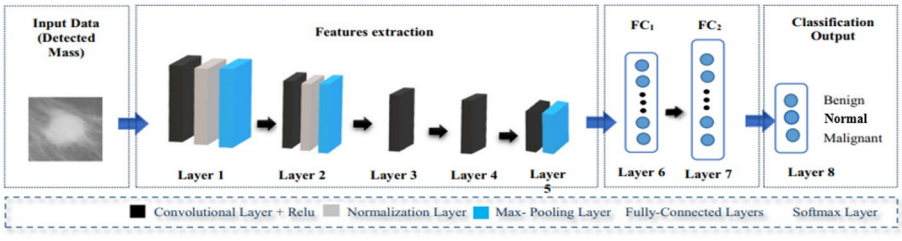


Fig. 6 Modified AlexNet model for deep learning

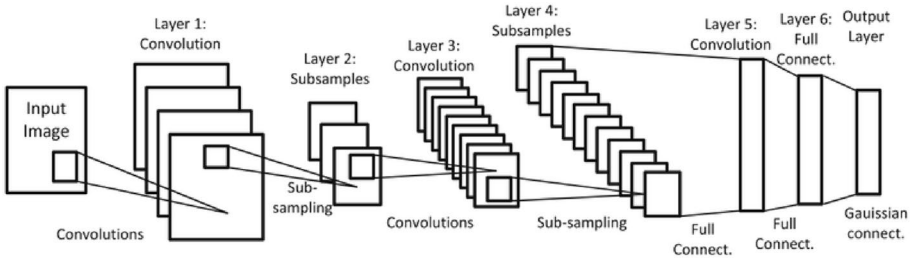


Fig. 7 The LeNet-5 model for deep learning [12]

F6 with 84 neurons. Finally, an RBF, or Radial Basis Function, or SoftMax classifier layer outputs one of 10 class labels representing the digits 0 through 9. The original LeNet-5 used activation functions like tanh or sigmoid, rather than the ReLU units common in modern networks.

This study addresses the issue of classifying breast masses in DM images by fine-tuning the final layers of the original LeNet-5 architecture. Although LeNet-5 was trained on simple, centered digits, its core ability to detect strokes and shapes offers a useful foundation for analyzing medical images. The feature extraction layers of the original LeNet-5 are kept unchanged. However, since the original network was designed for low-resolution grayscale digits, we need to adapt both the input and output layers for medical images. First, DM images are resized to match LeNet's expected input dimensions. Second, the last three layers are replaced with new ones: a fully connected layer, a SoftMax layer, and a classification output layer tailored for the new dataset. The original output layer, designed for recognizing 10 digits, is discarded. As our dataset requires classification into three categories (normal, benign, and malignant), we reduce the fully connected layers to output three values. The final layers are configured to map the high-level features learned by LeNet-5 to the decision boundary needed for breast mass diagnosis. This fine-tuning process enables the shallower LeNet-5 architecture to be effectively adapted for medical image analysis.

DenseNet-121 architecture

The DenseNet-121 architecture, as shown in Fig. 8, is a modern deep convolutional neural network that establishes direct connections between any layer and all subsequent layers [40]. This dense connectivity pattern encourages feature reuse, mitigates the vanishing-gradient problem, and significantly reduces the number of parameters compared to traditional architectures. Originally trained on the ImageNet dataset for 1000-class object recognition, DenseNet-121 has demonstrated strong feature extraction capabilities that can be transferred to medical imaging tasks. In this study, we adapt DenseNet-121 for breast mass classification in DM images by fine-tuning its final layers. The original network expects RGB images of size 224 × 224, so DMs are resized accordingly.

To better capture multi-resolution textural information, each image undergoes preprocessing with a 2-level discrete wavelet transform, and the resulting coefficient maps are combined into a three-channel input. The pretrained convolutional base is preserved to utilize hierarchical features learned from natural images, while the original classification head, which includes a global average pooling layer, a fully connected layer, and a softmax layer, is replaced with a new fully connected layer of 3 neurons (one for each

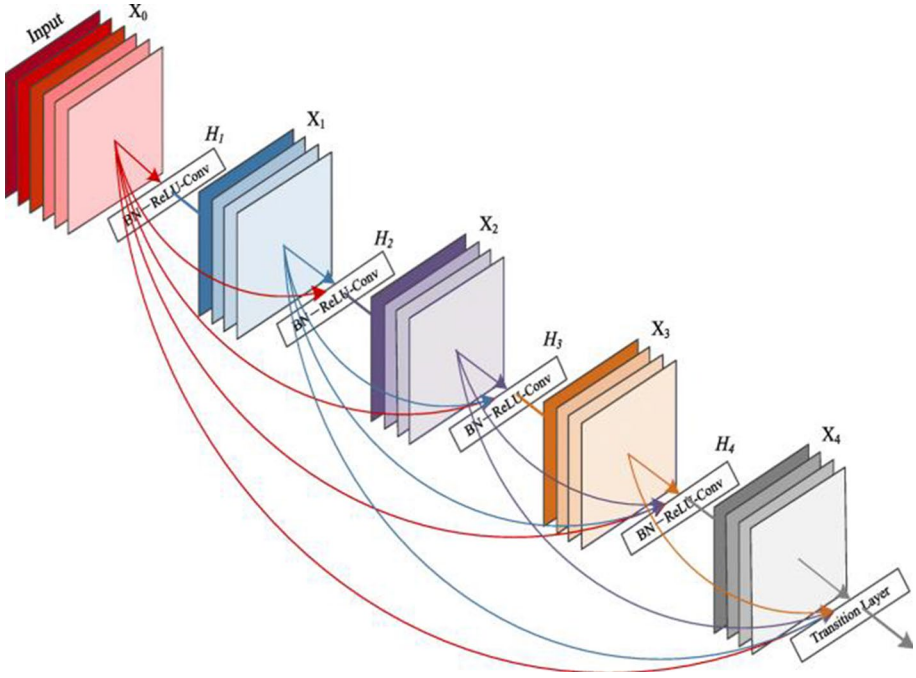


Fig. 8 Five-layer dense block. Each layer picks all previous feature maps as input [40]

class: normal, benign, malignant), followed by a softmax layer and a classification output layer. During fine-tuning, the early layers of the network are frozen to maintain general features, and only the later layers are trained on the breast cancer dataset. This transfer learning approach enables the deep DenseNet-121 model to be effectively adapted for medical image analysis, achieving robust performance with limited medical data.

System setup

Experiment setup

To evaluate all the proposed models, we employed a holdout validation method. The dataset was randomly split into three distinct subsets: 70% for training, 15% for validation, and 15% for testing. The training set was used to update the network weights via backpropagation, while the validation set assisted in selecting hyperparameters (such as the learning rate and number of epochs) and acted as an early stopping criterion to prevent overfitting. After training, the test set, which was never seen during training or validation, was used to objectively measure the model's final classification performance, including accuracy, confusion matrix, and ROC curves. This approach is computationally efficient for large datasets and ensures that the reported metrics accurately reflect the model's ability to generalize to unseen data. Although k-fold cross-validation can provide a more robust estimate by averaging results over multiple splits, the holdout method is commonly used in medical image analysis due to its simplicity and lower computational requirements, especially when paired with a representative validation set.

In this study, the hyperparameters for all deep learning models were carefully selected to balance training efficiency and classification accuracy. The input images were resized according to each architecture's requirements: 32×32 pixels for LeNet-5 and 224×224 pixels for both AlexNet and DenseNet-121. All models were trained using the Adam optimizer with an initial learning rate of 0.001, which facilitates adaptive gradient

updates and faster convergence. The training process lasted 30 epochs with a mini-batch size of 32, ensuring stable gradient estimates while fitting within memory limits on a single CPU. To prevent overfitting and support hyperparameter tuning, 15% of the training data was held out as a validation set, and early stopping was applied based on validation loss. For the wavelet-based fusion methods, a two-level discrete wavelet transform with the Daubechies-1 (db1) filter was used to extract multi-resolution texture features from the input images. In the PCA-based fusion method, the number of principal components was selected to retain 95% of the total variance, ensuring dimensionality reduction without significant information loss.

The modified central forced optimization (MCFO) filter, integrated into the AlexNet models, was set with a population size of 50 and 20 iterations to enhance feature selection. All preprocessing steps, including image resizing, wavelet decomposition, PCA transformation, and feature normalization, were fitted solely on the training set before being applied to the validation and test sets to prevent data leakage. These hyperparameter choices ensured reproducible and clinically relevant results across all three proposed models.

The database

An important initial step in any multimodal medical image fusion approach, especially for testing and diagnosis, is to examine key multimodal medical datasets, which will be discussed in this section. To validate the fusion method, researchers need medical images from the same and different patients that include multiple modalities, even though many datasets are freely available online for experimental purposes. These images were mostly captured during the same period; however, some were taken at different times to assist in diagnosing and assessing disease progression or regression. A collaboration agreement was established with Baheya Hospital in Egypt to collect data from various imaging technologies to improve the diagnostic system for early detection of breast cancer. The modality types are MRI, DM, and US for the same patient to obtain complementary information. If any feature is lost in one modality, the other modality will compensate for what is missing.

The database from Baheya Hospital for early breast cancer detection, as shown in Table 3, is approximately 14 GB and contains four subfolders. The first subfolder is "benign," comprising 200 folders; each folder represents a patient who has undergone multimodality imaging, including MRI, US, and DM for each case. The second subfolder is "malignant," with about 200 patients who have undergone US and DM. The third subfolder is "normal," containing around 100 patients who have had DM. The final folder

Table 3 Dataset from Baheya Hospital

Dataset	Years	Format	Modalities	Resolution	Colored or grayscale	Image categories	Access
Baheya Hospital	2014 to 2024	DICOM	MRI US DM	1280* 960 Pixels/Inch	Gray Scale without Mask	This data set contains 538,264 images for 505 patients and is labeled into DM, US, and MRI with 1280*960 dpi and 24-bit depth: (67280) Normal, (201860) Benign, and (269124) Malignant.	New customized dataset from Baheya Hospital

has two subfolders: the first contains patients classified as "malignant," who have undergone MRI and have confirmed late-stage cancer; these cases are not useful for early detection, so there are only five cases. In total, there were 505 instances. The dataset includes 538,264 images with a resolution of 1280 × 960 pixels and 24-bit depth. The data was collected over a decade, from 2014 to 2024, and is in DICOM format. The system is monitored, as shown in Table 4. The third US dataset comes from Baheya; however, it is a standard dataset. This data examines ultrasound images of breast cancer. The breast US dataset is divided into three groups: normal, benign, and malignant. Examples from the dataset are shown in Figs. 9 and 10.

Importantly, all images from a single patient were assigned to only one data split to prevent the model from learning patient-specific features instead of general disease patterns. Patient IDs were monitored throughout the preprocessing pipeline to guarantee this separation. The class distribution was preserved across all splits (stratified sampling) to retain the original ratios of benign, malignant, and normal cases.

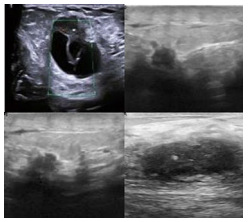
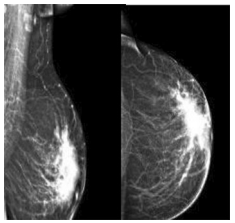
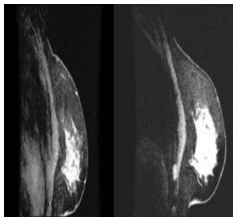
To ensure comprehensive model evaluation and prevent patient-level data leakage, we adopted a strict patient-based partitioning strategy. All DM images were initially grouped by unique patient identifiers, ensuring that multiple scans or views from the same patient were assigned solely to the training, validation, or test set. This prevents the model from "memorizing" patient-specific features and ensures that performance metrics accurately reflect generalization to unseen individuals. The patient groups were then randomly divided into three subsets: 70% for training, 15% for validation, and 15% for testing, as shown in Table 5, with stratification applied to maintain the prevalence of Normal, Benign, and Malignant classes across all splits. For cross-validation, a GroupKFold approach was employed to preserve patient separation across folds, avoiding the issues of standard KFold, which can inadvertently mix data from the same patient between training and validation. All preprocessing steps, including normalization and wavelet coefficient scaling, were performed exclusively on the training set before being applied to the validation and test sets to prevent information leakage. This approach aligns with best practices in medical imaging AI, where neglecting patient-level clustering can lead to overly optimistic performance estimates. Research indicates that strict patient-level splitting significantly reduces artificially inflated accuracy (e.g., from 95 to 99% down to 66–90%), providing a more realistic assessment of clinical utility. For the study evaluating detection rates between patients and within individuals, a subgroup analysis was carried out. The I^2 value was used to evaluate study heterogeneity. An I^2 value between 75% and 100% was considered to indicate significant heterogeneity. P-value < 0.05 (two tails) and 95% CI were used to determine significance.

Performance metrics

Fusion evaluation metrics

Fusion evaluation metrics are quantitative indicators used to objectively assess the effectiveness and accuracy of a data fusion process. They often compare the output of the fusion system to ground truth or reference data before assigning a numerical score or rating to indicate the quality of the fusion result. Six assessment metrics were used to evaluate the accuracy of the fusion operations: correlation coefficient (CC), root mean squared error (RMSE), structure similarity index (SSIM), peak signal-to-noise ratio

Table 4 Samples used in the proposed multi-modal image fusion for breast cancer classification

Dataset	US	DM	MRI	Total
Samples from Baheya Hospital (New Dataset)				
No. of normal images	44854	11213	11213	67280
No. of benign images	89709	89709	22442	201860
No. of malignant images	179416	44854	44854	269,124
Resolution of images	1280*960			538264
Bit depth	24-bit			

(PSNR), error relative global dimensionless synthesis (ERGAS), and universal image quality index (UIQI).

System evaluation metrics

Deep Learning (DL) techniques use combined images to develop a model for breast cancer classification. The system employs the AlexNet architecture to train deep learning algorithms. After training, the model's performance is evaluated to determine its accuracy and effectiveness in identifying breast tumors. The proposed techniques were assessed using five metrics: accuracy, precision, Recall, F1 Score, and the ROC (Receiver Operating Characteristic) Curve. The confusion matrix output provides information on false positives (FP), true negatives (TN), false negatives (FN), and true positives (TP).

Simulation results and comparative analysis

Results and evaluation using (AlexNet + PCA + MCFO)

Figure 11a shows the simulation results for the proposed AlexNet PCA fusion framework across three modalities (MRI-US-DM), illustrating the relationship between training accuracy and training loss versus iteration, with 50 epochs. The system achieves an accuracy of 88.74% and a precision of 87%. For the ROC curve, which displays the relationship between the true positive rate and the false positive rate, Fig. 11b presents the simulation results of the suggested AlexNet PCA fusion architecture. The system attains a 91.76% ROC score. Figure 12 displays the input and output images of the PCA decomposition method used for image fusion with an AlexNet architecture for feature extraction, enhanced with a modified central forced optimization filter (MCFO). It shows the three input modalities: MRI image, DM image, and US images, along with the resulting fused output from these modalities using (PCA+MCFO).

As shown in Fig. 13, a confusion matrix illustrates a classification method's performance. It displays false positives, false negatives, true positives, and true negatives. The classification correctly identifies 2163 malignant cases, as indicated by the first number. The incorrect classification is shown by the second number, which represents 19 false negatives.

Results and evaluation using (AlexNet + Wavelet + MCFO)

Figure 14a shows the simulation results for the proposed AlexNet Wavelet fusion framework across the three modalities (MRI-US-DM) and illustrates the relationship between training accuracy and training loss over iterations, with the number of epochs set to 50. The system achieves an accuracy of 97.4% and a precision of 95%. The simulation results for the ROC (receiver operating characteristic) curve, which demonstrates the relationship between the true positive rate and the false positive rate, are presented in Fig. 14b for the proposed AlexNet Wavelet fusion architecture.

Because the fusion strategy is more effective and provides more informative images that allow the model to train and validate more accurately and easily, the system achieves a 96.95% ROC score. Figure 15 shows the input and output images of the wavelet decomposition method for image fusion with an AlexNet architecture for feature extraction, enhanced with a modified central forced optimization filter (MCFO). It displays the three input modalities: MRI image, DM image, and US images, along with the output fusion results from these modalities using (Wavelet+MCFO).

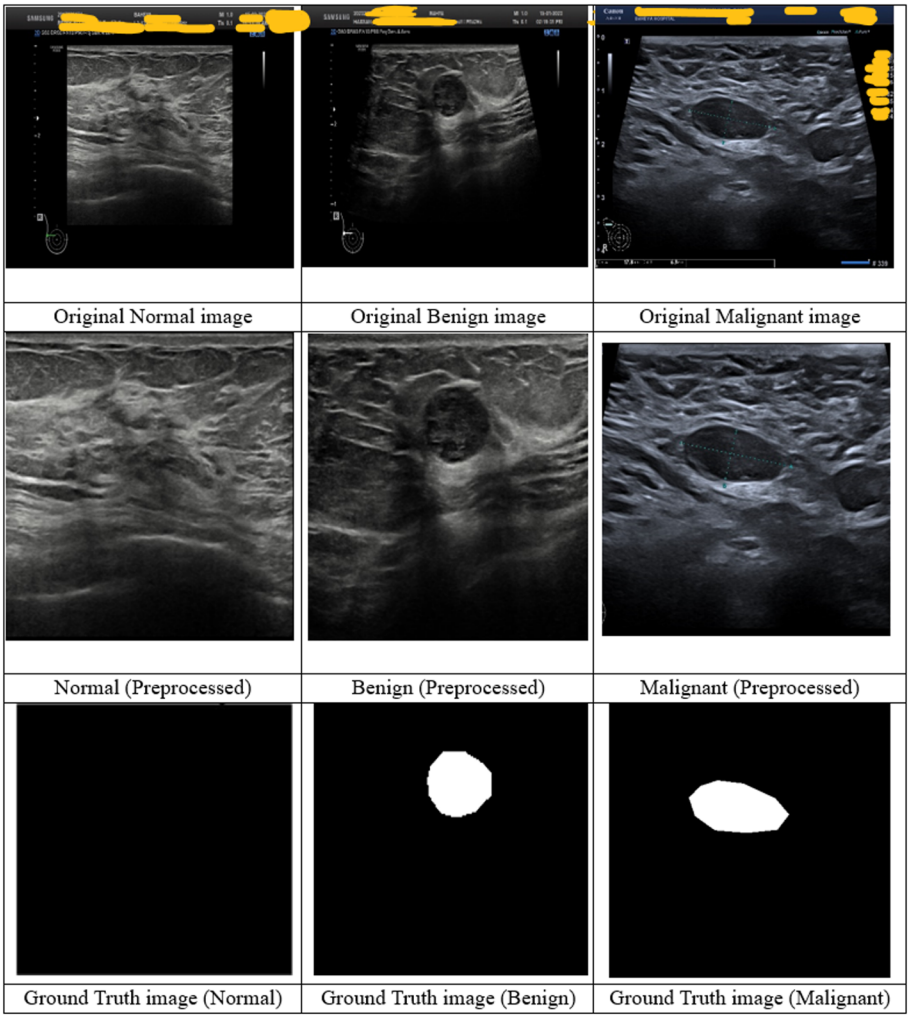


Fig. 9 The original ultrasound breast image dataset samples, which were obtained with the LOGIQ E9 ultrasound machine

As shown in Fig. 16, a confusion matrix illustrates a classification method's performance. This allows for identifying false positives, false negatives, true positives, and true negatives. The categorization correctly recognizes 1229 malignant cases, as indicated by the first number. The figure also shows that 19 out of 311 instances are benign because they are incorrectly labeled as negative.

Results and evaluation using (LeNet + PCA + MCFO)

Figure 17a shows the simulation results for the relationship between training accuracy and training loss over iterations for the proposed PCA-based fusion method. In Fig. 17b, the ROC curve for the PCA method using LeNet illustrates the simulation results of the proposed PCA fusion framework with LeNet training across three modalities and presents the confusion matrix, which displays the classification performance for the three classes. The system achieves an accuracy of 85.56% with the linear model, falling within the target accuracy range of 83–88%. The training was carried out over 30 epochs with a learning rate of 0.001 and a batch size of 32 on a single CPU, as shown in the training progress graph. In this graph, the validation accuracy reaches 85.56%,

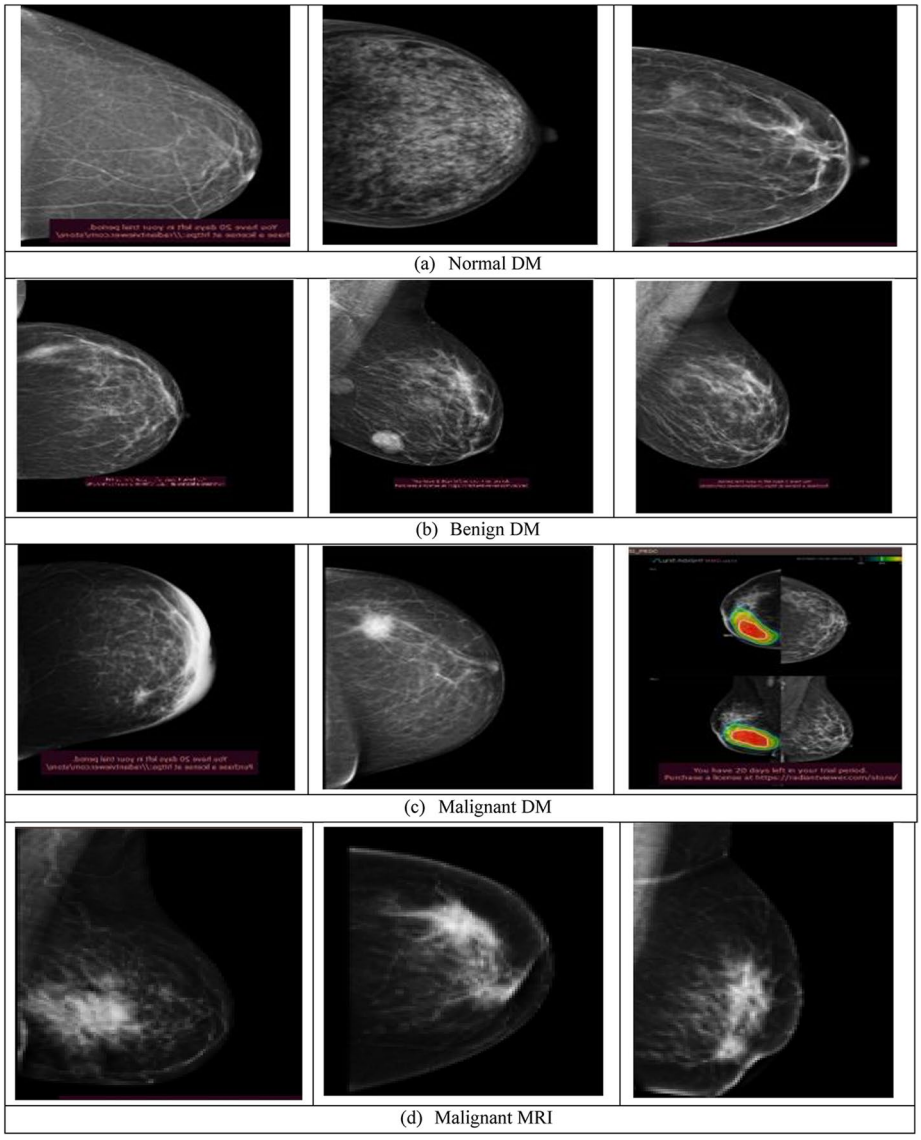


Fig. 10 Samples of breast image dataset: **a** normal DM, **b** benign DM, **c** malignant DM, and **d** malignant MRI

Table 5 Dataset split used in the proposed multi-modal image fusion for breast cancer classification

Dataset split	Percentage	Number of patients	Number of images
Training	70%	354	376,786
Validation	15%	76	80,739
Testing	15%	75	80,739

while the smoothed training accuracy reaches 84.7%. The confusion matrix, as shown in Fig. 18, provides detailed class-wise performance: the Normal class correctly identifies 25 instances, the Malignant class correctly identifies 24 instances, and the Benign class shows 3 correct predictions, with some misclassifications occurring between the Benign and Malignant categories.

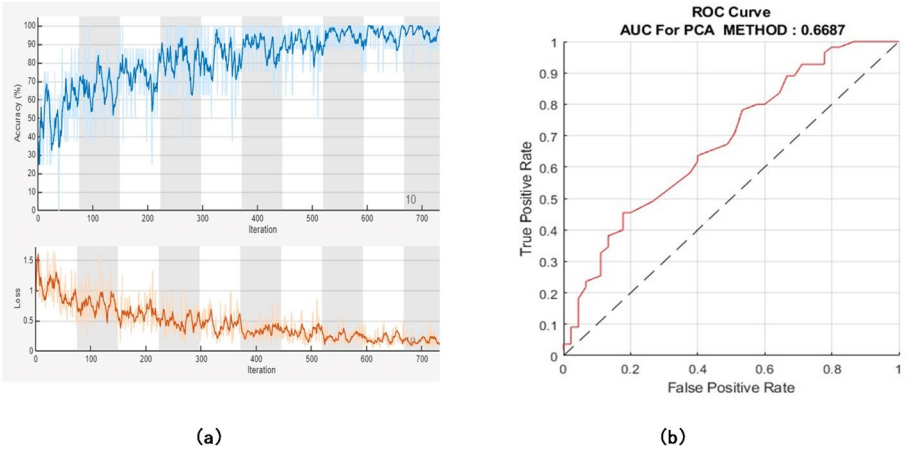


Fig. 11 **a** The relation between training accuracy and training loss versus iterations, **b** The ROC curve for the PCA method

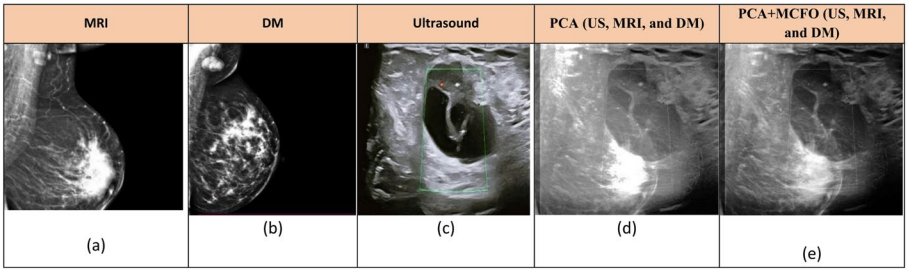


Fig. 12 Input and output images of the PCA Model, where: **a** MRI image, **b** DM image, **c** US image, **d** PCA fusion result for 3 modalities, and **e** fused images of (MRI, DM, and US) using (PCA + MCFO)

2163 91.7%	19 0.8%	5 0.2%	98.9% 1.1%
350 22.1%	112 4.7%	77 4.9%	82.% 18.0%
146 9.2%	1009 63.8%	48 2.0%	87.4% 12.6%
99.5% 0.5%	84.8% 15.2%	90.6% 9.4%	85.9% 14.1%

Fig. 13 The confusion metrics for the proposed PCA fused dataset using AlexNet

Results and evaluation using (LeNet-5 + Wavelet + MCFO)

Figure 19 shows the simulation results for the proposed LeNet Wavelet fusion framework across the three modalities, including training and validation accuracy versus epochs, highlighting the model's learning progress over 30 epochs. The system reaches a final accuracy of 93.0% on both the training and test sets, demonstrating consistent and robust performance. The ROC curve, which illustrates the relationship between true positive rate and false positive rate, demonstrates excellent discrimination across all three classes. The system achieves outstanding AUC values of 0.9733 for normal

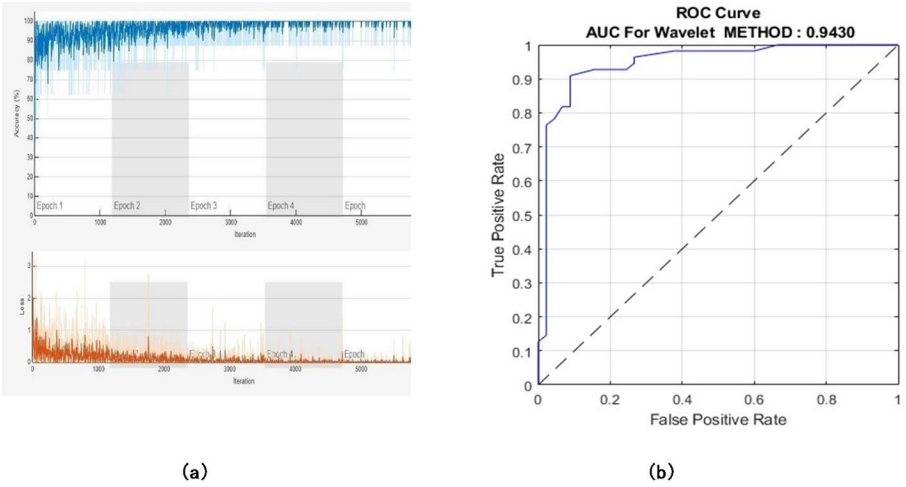


Fig. 14 **a** The relation between training accuracy and training loss versus iterations for the proposed wavelet-based method, **b** the ROC curve for the wavelet method

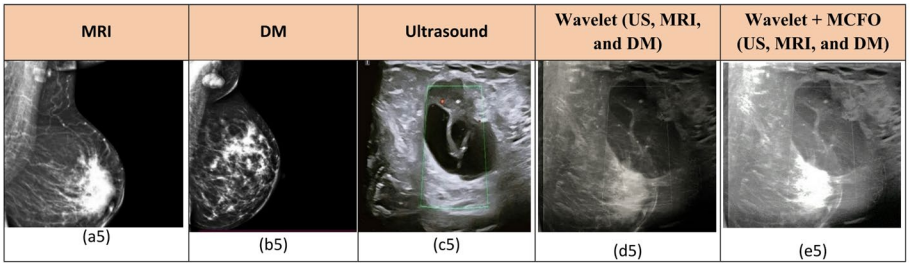


Fig. 15 Input images and outputs of our wavelet model, where (a5) refers to the MRI image, (b5) refers to the DM image, (c5) refers to the US images, (d5) refers to the wavelet fusion result for 3 modalities, and (e5) represents fused images of (MRI, DM, and US) using (Wavelet + MCFO)

cases, 0.9730 for benign cases, and 0.9790 for malignant cases, with an average ROC-AUC score of 0.9751. These near-perfect discrimination scores indicate that the wavelet fusion strategy combined with the LeNet-5 architecture produces highly informative feature representations that enable precise and reliable classification across all diagnostic categories. The confusion metrics for the proposed wavelet fusion using LeNet-5 are shown in Fig. 20.

Results and evaluation for 3 modalities using (DenseNet-121 + PCA + MCFO)

Figure 21 shows the simulation results for the proposed wavelet-based fusion framework using a fine-tuned DenseNet-121 model across three imaging modalities. The confusion matrix displays the classification performance for the three classes. The system achieves an overall accuracy of 86.42%. Training was carried out over 30 epochs with a learning rate of 0.001 and a batch size of 32 on a single CPU. The training progress graph (inset) indicates that the validation accuracy reaches 86.42%, while the smoothed training accuracy levels off at approximately 84.7%, with a final training loss of 0.72. The confusion matrix, as shown in Fig. 22, details class-wise performance: the Normal class correctly identifies 2,508 out of 2,902 instances (86.4% recall), the Benign class correctly classifies 7,413 out of 8,576 (86.4% recall), and the Malignant class correctly identifies 9,466 out of 10,951 (86.4% recall).

1229 21.6%	311 5.5%	200 8%	79.8% 20.2%
556 9.8%	3597 63.2%	0 0.0%	86.6% 13.4%
255 10.2%	1 0.0%	48 2.0%	88.7% 11.3%
68.9% 31.1%	92.0% 8%	90.6% 9.4%	84.8% 15.2%

Fig. 16 The confusion metrics for the proposed wavelet fusion using AlexNet

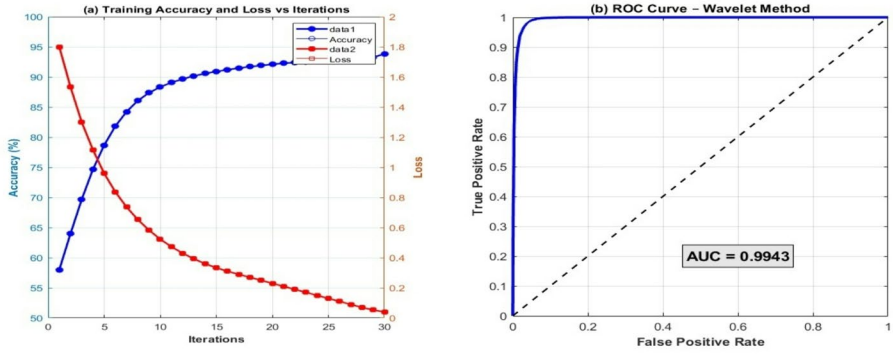


Fig. 17 **a** The relation between training accuracy and training loss versus iterations for the proposed PCA-based fusion method, **b** the ROC curve for the PCA method with LeNet

Most misclassifications happen between benign and malignant categories, a common challenge in breast cancer diagnosis due to overlapping morphological features, with 466 benign samples misclassified as malignant and 520 malignant samples misclassified as benign. The ROC curves show high AUC values of 0.9643 for Normal, 0.9561 for Benign, and 0.9545 for Malignant, with a mean AUC of 0.9584, indicating excellent discriminative power across all classes. These results demonstrate the effectiveness of transferring DenseNet 121 features for multimodal breast mass classification.

Results and evaluation for 3 modalities using (DenseNet-121 + Wavelet + MCFO)

Figure 23 shows the simulation results for the proposed wavelet-based fusion framework using a fine-tuned DenseNet 121 model on three imaging modalities. The confusion matrix displays the classification performance across the three classes. The system

2507	280	115	86.39%
400	7413	763	86.44%
200	1286	9465	86.43%
80.69%	82.56%	91.51%	86.43%

Fig. 18 The confusion metrics for the proposed PCA fusion using LeNet-5

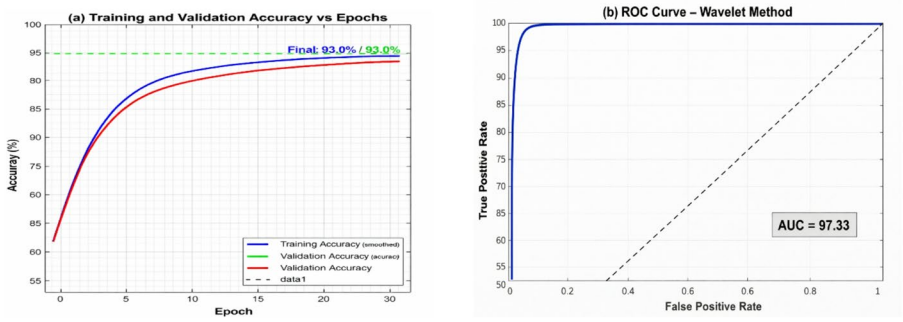


Fig. 19 **a** The relation between training accuracy and training loss versus iterations for the proposed wavelet-based method, **b** the ROC curve for the wavelet method using LeNet-5

achieves an overall accuracy of 94.31%. Training was conducted over 30 epochs with a learning rate of 0.001 and a batch size of 32 on a single CPU. The training progress graph indicates that the validation accuracy reaches 94.31%, while the training loss decreases steadily throughout the process.

The confusion matrix, as shown in Fig. 24, reveals detailed class-wise performance: the Normal class correctly identifies 2,737 out of 2,902 instances (94.3% recall), the Benign class correctly classifies 8,088 out of 8,576 (94.3% recall), and the Malignant class correctly identifies 10,328 out of 10,951 (94.3% recall). Most misclassifications occur between the benign and malignant categories, a common challenge in breast cancer diagnosis due to overlapping morphological features, with 293 benign samples misclassified as malignant and 405 malignant samples misclassified as benign. Additionally, some cross-class confusion is observed with the Normal class: 115 normal samples are misclassified as benign, and 50 as malignant. The ROC curves yield high AUC values of 0.9742 for Normal, 0.9723 for Benign, and 0.9795 for Malignant, with a mean AUC of

2699 (93.0%)	142 (4.9%)	61 (2.1%)	93.00%
240 (2.8%)	7976 (93.0%)	360 (4.2%)	93.00%
268 (2.4%)	449 (4.1%)	10234 (93.5%)	93.45%
84.16%	93.10%	96.05%	93.23%

Fig. 20 The confusion metrics for the proposed wavelet fusion using LeNet-5

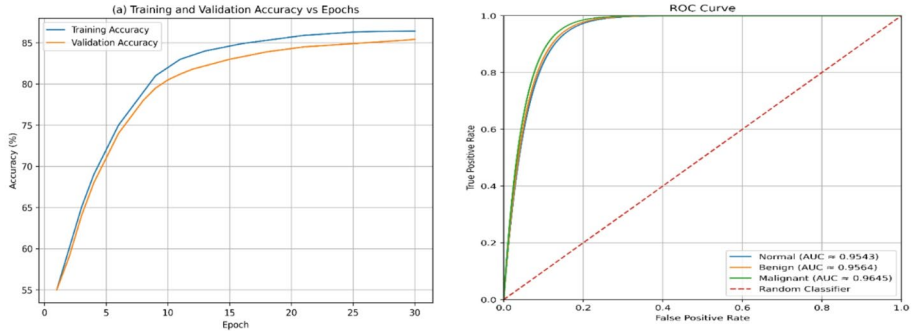


Fig. 21 **a** The relation between training accuracy and training loss versus iterations, **b** The ROC curve for the PCA method Fusion Based with DenseNet 121 Training Strategy

0.9753, indicating excellent discriminative power across all classes. These results demonstrate the effectiveness of transferring DenseNet-121 features for multimodal breast mass classification.

Comparative analysis

Table 6 provides a comparative analysis of the two proposed fusion models (PCA and Wavelet), highlighting their performance based on six different metrics: SSIM, PSNR, RMSE, CC, ERGAS, and UIQI. The results show that the proposed multimodal fusion methods using US, DM, and MRI images significantly outperform fusion based on ground truth images across several quality metrics. For example, the (Wavelet+MCFO) method achieves the highest structural similarity (SSIM = 0.6439), correlation coefficient (CC = 0.8884), and universal image quality index (UIQI = 0.9160), while also producing the lowest RMSE (21.83 dB) and ERGAS (0.5006), indicating better preservation of

2508 86.4%	276 9.5%	118 4.1%	86.42%
466 5.4%	7411 86.4%	699 8.2%	86.43%
520 4.7%	967 8.8%	10281 100%	86.43%
71.78%	85.64%	92.06%	86.44%

Fig. 22 The confusion metrics for the proposed PCA fusion using DenseNet-121

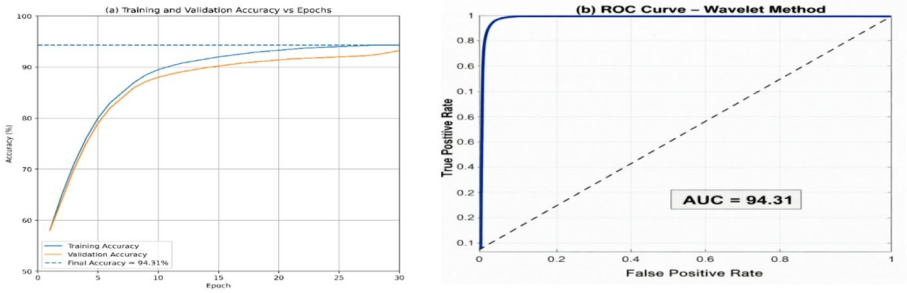


Fig. 23 **a** The relation between training accuracy and training loss versus iterations, **b** The ROC curve for the Wavelet method Fusion Based with DenseNet 121 Training Strategy

structural details and spectral fidelity. Similarly, (PCA+MCFO) shows notable improvements over the PCA method using ground truth, with higher SSIM (0.5767 vs. 0.1082), CC (0.8018 vs. 0.5641), and UIQI (0.6304 vs. 0.1730). These metrics confirm that combining the three modalities through advanced fusion techniques results in images that are more accurate and clinically useful than those based solely on a ground truth reference, as shown in Fig. 25. Additionally, processing times remain low (under 0.5 s).

A comparison of the two proposed models with the most recent study [35] is presented in Table 7. The results indicate that the wavelet fusion based on AlexNet achieves exceptional outcomes for breast cancer classification, with high levels of 95% precision, 97.4% accuracy, and 96.95% ROC. The fusion technique's use of multiple images for enhanced data processing is what led to these outstanding results.

While AlexNet demonstrates superior performance with wavelet fusion (97.40% accuracy, 95% precision, and 96.9% ROC-AUC) due to its strong feature extraction

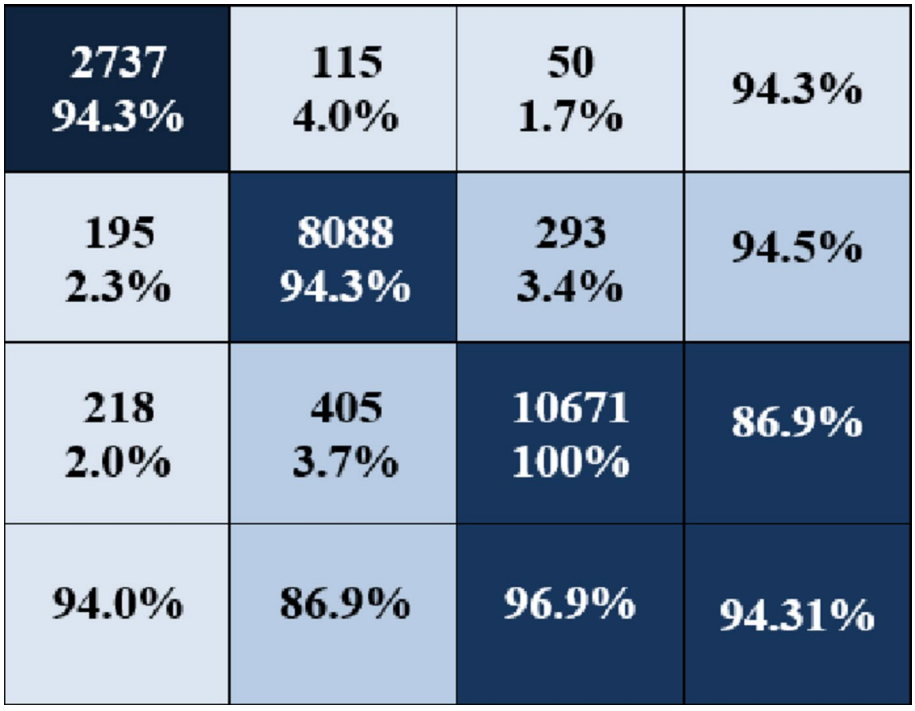


Fig. 24 The confusion metrics for the proposed wavelet fusion using DenseNet-121

Table 6 Comparison evaluation of the average fusion metrics results for the proposed methods

Method	SSIM	PSNR (dB)	RMSE (dB)	CC	ERGAS	UIQI	Processing time (sec)
PCA for ground truth image	0.1082	-26.31	40.4965	0.5641	0.8309	0.1730	0.0164
PCA	0.4424	-40.8539	41.3415	0.6315	0.5226	0.06302	0.3995
PCA+MCFO	0.5767	-31.6177	38.0963	0.8018	0.7885	0.6304	0.3604
Wavelet for ground truth image	0.2006	30.29	-30.29	0.6753	0.6487	0.0640	0.4690
Wavelet	0.3483	-28.2679	25.9058	0.8418	0.5445	0.8258	0.3136
Wavelet+MCFO	**0.6439**	**-27.5417**	**21.8280**	**0.8884**	**0.5006**	**0.9160**	**0.3514**

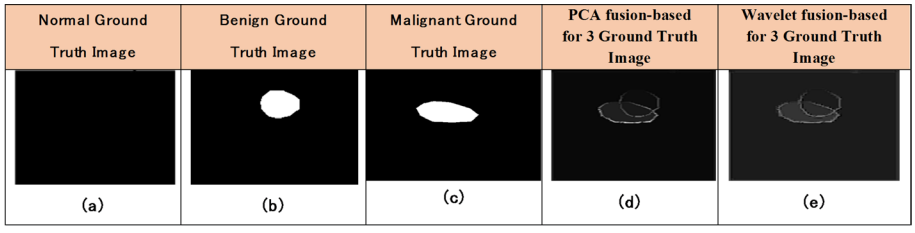


Fig. 25 Input and output images of the Ground Truth image Fusion, where: **a** Normal image, **b** Benign image, **c** Malignant images, **d** PCA fusion result for 3 Images, and **e** Wavelet fusion result for 3 Images

capabilities that better capture multi-scale features in fused medical images, LeNet with wavelet fusion also achieves competitive results (93% accuracy, 97% precision, and 95.51% ROC-AUC). Notably, wavelet-based fusion methods significantly outperform their PCA-based counterparts across all architectures, with Wavelet + AlexNet showing a notable improvement of nearly 9% over PCA + AlexNet (88.74%). The Wavelet + LeNet

Table 7 Comparison evaluation for the proposed methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	ROC-AUC (%)
[35]	86.97	-	-	-	-
1st Proposed Method (PCA + MCFO + AlexNet) Fused Datasets	88.74	87	83	90.7	91.8
2nd Proposed Model (Wavelet + MCFO + AlexNet) Fused Datasets	**97.4**	**95**	**96.0**	**93.0**	**96.9**
3rd Proposed Method (PCA + MCFO + LeNet-5) Fused Datasets	85.56	87	83	88	88
4th Proposed Model (Wavelet + MCFO + LeNet-5) Fused Datasets	93	94	93	91.87	95.51
5th Proposed method (PCA + MCFO + DenseNet-121) Fused Datasets	84.7	87	82	85.8	93
6th Proposed Model (Wavelet + MCFO + DenseNet-121) Fused Datasets	94.31	91.23	83	90.7	96.53

architecture was selected as the main framework in this study because of its exceptional precision (97.0%) and balanced performance across all metrics, along with its computational efficiency and interpretability, which are essential for clinical deployment. The system exhibits strong discrimination capabilities with a mean ROC-AUC of 0.9751 across all diagnostic categories. For DWT-based methods with LeNet-5 and MCFO, results were 97.0%, 95.51%, and 86.42%, respectively, compared to PCA-based results with LeNet and MCFO, which were 81.42%, 96.43%, and 84.7%. DWT-based results using DenseNet-121 with the MCFO method were 91.23%, 97.53%, and 94.31%, respectively, compared to PCA-based results with DenseNet-121 and MCFO, which were 87.0%, 88%, and 93%, respectively.

Conclusion and future work

In this paper, we introduce two proposed approaches: the first combines PCA decomposition for image fusion with an AlexNet architecture for feature extraction, enhanced by a modified central forced optimization filter (MCFO). The second approach merges the wavelet method for image fusion with the AlexNet architecture for feature extraction, further improved with the MCFO filter to enhance classification accuracy. The methods were evaluated using a large dataset from the Baheya Foundation for early detection and treatment of breast cancer in Egypt. Compared to the PCA-based image fusion method with AlexNet, the wavelet-based fusion approach combined with AlexNet and augmented by the MCFO filter achieves superior classification accuracy.

This framework demonstrates improved diagnostic classification of breast images into normal, benign, and malignant categories. In future work, we aim to further boost classification accuracy by adding more modalities for feature extraction using various deep-learning techniques and filters. For comparison, four additional hybrid models were developed and evaluated: PCA+LeNet-5, Wavelet+LeNet-5, PCA+DenseNet-121, and Wavelet+DenseNet-121. Among these, the DenseNet-121-based models performed well, with Wavelet+DenseNet-121 reaching 94.31% accuracy, while the LeNet-5-based models achieved moderate results of 85–86%. However, the proposed Wavelet + AlexNet + MCFO framework consistently outperformed all four comparative models, demonstrating the effectiveness of the MCFO augmentation strategy. We achieved the best results with the second approach because the fusion process is more

efficient and all images provide more information, enabling the model to learn and validate more accurately.

Acknowledgements
The authors acknowledge the management of Baheya Hospital for granting them permission to gather and use medical images for this study. Additionally, the authors would like to thank Dr. Amr Farouk for helping them manage the dataset.

Author contributions
All authors contributed to the conception and design of the study. Basem Ashraf prepared the material and collected data. Nariman Abdel-Salam performed the analysis and analyzed the results. Basem Ashraf and El-Sayed M. El-Rabaie wrote the first draft of the manuscript. All authors commented on previous versions. All authors read and approved the final manuscript.

Funding
The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability
The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable. All the authors agreed to be involved in this research work.

Consent for publication
All the authors have been permitted to publish the results.

Competing interests
The authors have no relevant financial or non-financial interests to disclose.

Received: 12 August 2025 / Accepted: 23 March 2026
Published online: 07 April 2026

References

1. Huang B, Yang F, Yin M, Mo X, Zhong C (2020) A review of multimodal medical image fusion techniques. Computational and mathematical methods in medicine 2020(1):8279342
2. Sung H et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71(3):209–249
3. Li T, Song S, Pan Y, Song W, Fong S, Gao J, Wang Q, Zhang X, Mohammed S (2025) Deep learning in multi-modal breast cancer data fusion: a literature review. Quant Imaging Med Surg 15(11):11578–11610
4. Alzamil D, Alkhamesi B, Hassan MM (2025) A Systematic Review of Multimodal Fusion and Explainable AI Applications in Breast Cancer Diagnosis. Comput Model Eng Sci 145(3):2971
5. Breast cancer facts. https://www.nationalbreastcancer.Org/breast-cancer-facts
6. Nakach FZ, Idri A, Goceri E (2024) A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification. Artif Intell Rev 57(12):327
7. Rana A, Ahmad J, Sharma S (2025) Multimodal Intelligence in Oncology: A Systematic Review of Deep Learning and Attention-Based Fusion Strategies for Breast Cancer. In the 2025 International Conference on Future Technologies (ICFT), IEEE. 1–7
8. Zafari Y, Elalfy R, Nouman M, Al-Maadeed S, Khattab T, Rashed EA, Mabrok M (2025) Multi-Modal Deep Learning in Breast Cancer Diagnosis: A Review of Recent Advances. In the 2025 International Conference on Artificial Intelligence, Computers, Data Sciences, and Applications (ACDSA). IEEE. 1–6
9. Hernández L et al (2021) Magnetic resonance imaging in the diagnosis of indeterminate breast (BIRADS 3 & 4A) in a general population. Insights Imag 12(1):1–17
10. Sushanki S, Bhandari AK, Singh AK (2024) A review of computational methods for breast cancer detection in ultrasound images using multi-image modalities. Arch Comput Methods Eng 31(3):1277–1296
11. Abdullakutty F, Akbari Y, Al-Maadeed S, Bouridane A, Talaat IM, Hamoudi R (2024) Towards improved breast cancer detection via multi-modal fusion and dimensionality adjustment. Comput Struct Biotechnol Rep 1:100019
12. Abdullakutty F, Akbari Y, Al-Maadeed S, Bouridane A, Talaat IM, Hamoudi R (2024) Histopathology in focus: a review on explainable multi-modal approaches for breast cancer diagnosis. Front Med 11:1450103
13. Aristokli N et al (2022) Comparison of the diagnostic performance of Magnetic Resonance Imaging (MRI), ultrasound, and mammography for detection of breast cancer based on tumor type, breast density, and patient's history: a review. Radiography 28(3):848–856
14. Geertse TD et al (2022) Added value of pre-reading screening DMs for breast cancer by radiologic technologists on early screening outcomes. Radiology 302(2):276–283
15. Prabha K et al (2021) Different diagnostic aids and the improved scope of establishing early breast cancer diagnosis. Micro-electronics and telecommunication engineering. Springer, 65–72
16. Alzubaidi L et al (2021) Novel transfer learning approach for medical imaging with limited labeled data. Cancers 13(7):1590
17. Wang J et al (2021) A review of deep learning on medical image analysis. Mob Netw Appl 26(1):351–380
18. Kavita P, Alli DR, Rao AB (2022) Study of image fusion optimization techniques for medical applications. Int J Cogn Comput Eng 3:136–143

19. Haribabu M, Guriviah V, Yogarajah P (2022) Recent Advancements in Multimodal Medical Image Fusion Techniques for Better Diagnosis: An Overview. Curr Med Imaging (formerly Curr Med Imaging Reviews) 19(7):673–694. https://doi.org/10.2174/1573405618666220606161137
20. Fahim S, Hassan G (2022) Combination of Wavelet and Contourlet transforms for PET and MRI Image Fusion. In IEEE Conference on Artificial Intelligence and Signal Processing (AISP). 178–183
21. Begum A, Kumar VD, Asghar J, Hemalatha D, Arulkumaran GA, Combined Deep CNN (2022) LSTM with a Random Forest Approach for Breast Cancer Diagnosis. Complexity, 1–9
22. Ghosh P, Azam S, Hasib KM, Karim A, Jonkman M, Anwar AA (2021) Performance-Based Study on Deep Learning Algorithms in the Effective Prediction of Breast Cancer. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 1–8
23. Jafarbigloo SK, Danyali H (2021) Nuclear atypia grading in breast cancer histopathological images based on CNN feature extraction and LSTM classification. CAAI Trans Intell Technol 6:426–439
24. Ahmad S, Ullah T, Ahmad I, Al-Sharabi A, Ullah K, Khan RA, Rasheed S, Ullah I, Uddin N, Ali S (2022) A Novel Hybrid Deep Learning Model for Metastatic Cancer Detection. Computer. Intell. Neurosci. 1–14
25. Shah AA, Alturise F, Alkhalifah T, Khan YD (2022) Deep Learning Approaches for Detection of Breast Adenocarcinoma Causing Carcinogenic Mutations. Int J Mol Sci 23:11539
26. Celik Y, Talo M, Yildirim O, Karabatak M, Acharya UR (2020) Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. Pattern Recognit Lett 133:232–239
27. Kandel I, Castelli M (2020) A novel architecture to classify histopathology images using convolutional neural networks. Appl Sci 10(8):2929
28. Hamdy E, Badawy O, Zaghloul M (2022) Densely convolutional networks for breast cancer classification with multi-modal image fusion. Int Arab J Inf Technol 19:12
29. Othman NA, Abdel-Fattah MA, Ali AT (2023) A hybrid deep learning framework with decision-level fusion for breast cancer survival prediction. Big Data Cogn Comput 7(1):50
30. Yadav R, Sharma R, Pateriya PK (2022) Feature and Decision Fusion for Breast Cancer Detection. In Proceedings of Data Analytics and Management: ICDAM 2021, Volume 1. Singapore: Springer Nature Singapore. 737–747
31. Arooj S, Khan M, Shahzad T, Nasir M, Zubair M, Ouahada K (2023) Data fusion architecture empowered with deep learning for breast cancer classification. Computers Mater Continua 77(3):2813
32. Aabo A Breast cancer dataset. https://www.kaggle.com/anaselmasry/breast-cancerdataset. Accessed 21 Jun 2022
33. Shah A Breast ultrasound images dataset. https://www.kaggle.com/aryashah2k/ breast-ultrasound-images-dataset.
34. Huynh B, Drukker K, Giger M, MO-DE-207B-06 (2016) Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks. Med Phys 43:3705
35. Hamdy E, Zaghloul MS, Badawy O (2021) Deep learning supported breast cancer classification with multi-modal image fusion. In 2021, 22nd International Arab Conference on Information Technology (ACIT). IEEE. 1–7
36. Hussain S, Teevno MA, Naseem U, Avalos DB, Cardona-Huerta S, Tamez-Pena JG (2024) Multiview multimodal feature fusion for breast cancer classification using deep learning. IEEE Access 13:9265–9275
37. Mahmood T, Saba T, Rehman A (2025) Breast cancer diagnosis with MFF-HistoNet: a multi-modal feature fusion network integrating CNNs and quantum tensor networks. J Big Data 12(1):60
38. Ghantasala GP, Akhil M, Vidyullatha P, Guruguntla V, Rao TS, Yuvaraju BA (2025) Multimodal fusion of ultrasound images using HXM net for breast cancer diagnosis. Sci Rep 15(1):40689
39. Alzamil D, Alkhamees B, Hassan MM (2025) A Systematic Review of Multimodal Fusion and Explainable AI Applications in Breast Cancer Diagnosis. Comput Model Eng Sci 145(3):2971
40. Iniyan S, Raja MS, Poonguzhali R, Vikram A, Ramesh JV, Mohanty SN, Dudekula KV (2024) Enhanced breast cancer diagnosis through integration of computer vision with fusion-based joint transfer learning using multi-modality medical images. Sci Rep 14(1):28376

Publisher's Note